ZooKeys 50: i-iv (2010)
doi: 10.3897/zookeys.50.543
www.pensoftonline.net/zookeys

EDITORIAL

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Taxonomy shifts up a gear: New publishing tools to accelerate biodiversity research

Biodiversity is under severe threat. It is estimated that there are untold millions of species on the planet and we have, in the past 250 years and recently with a decreasing number of specialists, described only about 2 million of them. Though this happens with an agonizing low description rate of an estimated number of 20,000 new species per year (Polaszek et al. 2005), digital and genomic resources now are allowing discovery of new and identification of existing species at a never before seen pace.

The principles of Open Access greatly facilitate dissemination of information through the Web where it is freely accessed, shared and updated in a form that is accessible to indexing and data mining engines using Web 2.0 technologies. Web 2.0 turns the taxonomic information into a global resource well beyond the taxonomic community. A significant bottleneck in naming species is the requirement by the current Codes of biological nomenclature ruling that new names and their associated descriptions must be published on paper, which can be slow, costly and render the new information difficult to find. In order to make progress in documenting the diversity of life, we must remove the publishing impediment in order to move taxonomy "from a cottage industry into a production line" (Lane et al. 2008), and to make best use of new technologies warranting the fastest and widest distribution of these new results.

In this special edition of ZooKeys we present a practical demonstration of such a process. The issue opens with a forum paper from Penev et al. (doi: 10.3897/zookeys.50.538) that presents the landscape of semantic tagging and text enhancements in taxonomy. It describes how the content of the manuscript is enriched by semantic tagging and marking up of four exemplar papers submitted to the publisher in three different ways: (i) written in Microsoft Word and submitted as non-tagged manuscript (Stoev et al., doi: 10.3897/zookeys.50.504); (ii) generated from Scratchpads (Blagoderov et al., doi: 10.3897/zookeys.50.506 and Brake and Tschirnhaus, doi: 10.3897/zookeys.50.505); (iii) generated from an author's database (Taekul et al., doi: 10.3897/zookeys.50.485). The latter two were submitted as XML-tagged manuscript. These examples demonstrate the suitability of the workflow to a range of possibilities that should encompass most current taxonomic efforts. To implement the aforementioned routes for XML mark up in prospective taxonomic publishing, a special software tool

Copyright *Lyubomir Penev et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

(Pensoft Mark Up Tool, PMT) was developed and its features were demonstrated in the current issue. The XML schema used was version #123 of TaxPub, an extension to the Document Type Definitions (DTD) of the US National Library of Medicine (NLM) (http://sourceforge.net/projects/taxpub/).

A second forum paper from Blagoderov et al. (doi: 10.3897/zookeys.50.539) sets out a workflow that describes the assembly of elements from a Scratchpad taxon page (http://scratchpads.eu) to export a structured XML file. The publisher receives the submission, automatically renders the file into the journal's layout style as a PDF and transmits it to a selection of referees, based on the key words in the manuscript and the publisher's database. Several steps, from the author's decision to submit the manuscript to final publication and dissemination, are automatic. A journal editor first spends time on the submission when the referees' reports are received, making the decision to publish, modify or reject the manuscript. If the decision is to publish, then PDF proofs are sent back to the author and, when verified, the paper is published both on paper and on-line, in PDF, HTML and XML formats. The original information is also preserved on the original Scratchpad where it may, in due course, be updated. A visitor arriving at the web site by tracing the original publication will be able to jump forward to the current version of the taxon page.

The exemplar papers are published in four different formats: (1) high-resolution, full-colour print version, to satisfy the current requirements of the International Code of Zoological Nomenclature (ICZN), as well as the readers who prefer hardcopy, and for the purposes of paper archiving; (2) PDF to provide an electronic version identical to the printed one, to be archived in BHL and PubMedCentral; (3) HTML to provide links to external resources and semantic enhancements to published texts for interactive reading, and (4) XML version based on the TaxPub XML schema to provide archiving document format for PubMedCentral and a machine-readable copy of the contents to facilitate future data mining.

Publication of the papers also triggers registration of the names in ZooBank, which will in due course be automated. In addition, the PDF versions are uploaded in the Biodiversity Heritage Library, the XML version, together with the PDF and separate files of figures into PubMedCentral and the taxon treatments are uploaded to Plazi. All the key elements are associated with doi numbers to ensure that they are persistently accessible.

The process described and demonstrated in this edition represent a degree of automation that is expected to reduce the cost of publication. More importantly it puts information into a form that is easily discovered, extracted and re-used. The most important result of this development is that we can expect to see a reduction in synonymy of new names, first because authors can easily check whether a name is currently in use, and second because the publication process itself is only a few weeks, so the chances of concurrent discovery of a new taxon are reduced. Another important consideration is that authors can publish large, monographic revisions incrementally, delivering one or a few taxonomic acts at a time, rather than taking years to produce a major monograph.

Special attention was paid to the semantic enhancements of the published texts. For the first time, a newly published taxonomic revision can be searched and retrieved for taxon treatments. For instance, one can map geographical coordinates either from the whole paper or in separate species treatments or for a genus treatment encompassing several species. The Pensoft Taxon Profile (PTP) is a new tool that provides dynamic extraction and display of information from selected web resources just by clicking on any taxon name mentioned in the publication. The tool is similar to existing dynamic content builders such as ispecies.org, but it encourages direct dynamic exploration as part of the reading process. The tool's "Create your own taxon profile" function can be used by the reader to explore taxa not mentioned in the paper. Literature citations, and materials in figures and tables are cross-linked through the text to external resources whenever possible.

Of special importance in revolutionising the way published taxonomic information is being disseminated and archived are the tools developed to provide automated export of the XML content to be harvested by large international organisations, particularly Encyclopedia of Life and Plazi. The system allows the information to be harvested, indexed and archived on the day of publication.

ZooKeys thus becomes the first taxonomic journal to provide a complete XML-based editorial, publication and dissemination workflow implemented as routine. Moreover, the same process is implemented not only at article level but also at taxon treatment level, which provides "atomisation" of a taxonomic publication into sections that may be harvested separately from the whole text and allows linking with a large number of external resources through the oldest existing identifier in biology, the taxon name! The same advanced workflow will also soon be implemented in botany through *PhytoKeys*, a forthcoming partner journal of *ZooKeys*. The semantic markup and enhancements are expected to greatly extend and accelerate the way taxonomic information is published, disseminated and used. The editors consider that this represents an important step in the "industrialisation" of taxonomy and a move firmly away from the unsustainable "business as usual" mindset. Needless to say, the editors and authors involved in the current issue are thrilled by this development and would like to receive comments and critisim for further developing of the proposed workflow.

Lyubomir Penev, David Roberts, Vincent Smith, Donat Agosti, Terry Erwin
Sofia – London – Tehran – Washington, 30th of June 2010

References

Lane R, Ashburner M, Ausubel JH, Blaxter M, Costello M, Dayrat B, Donoghue M, Edwards J, Hirsch L, Le Gal, L, Godfray C, Johnson K, Knapp S, Krishtalka L, Kuntner M, May R, McNeely J, Remsen D, Smith R, Tillier S, Wägele W (2008) Taxonomy in Europe

in the 21st century. Report to the Board of Directors European Distributed Institute of Taxonomy. http://ww2.bgbm.org/EditDocumentRepository/Taxonomy21report.pdf

Polaszek A, Agosti D, Alonso-Zarazaga M, Beccaloni G, Bjørn PdP, Bouchet P, Brothers DJ, Cranbrook G, Evenhuis NL, Godfray HCJ, Johnson NF, Krell F-T, Lipscom D, Lyal CHC, Mace GM, Mawatari S, Miller SE, Minelli A, Morris S, Ng PKL, Patterson DJ, Pyle RL, Robinson NJ, Rogo L, Taverne J, Th ompson FC, Tol J van, Wheeler QD, Wilson EO (2005) A universal register for animal names. Nature 437: 4. doi: 10.1038/437477a